WHAT IS CLAIMED IS:

1. A method of using stereo vision to interface with a computer, the method comprising:

capturing a stereo image;

processing the stereo image to determine position information of an object in the stereo image, the object being controlled by a user; and

using the position information to allow the user to interact with a computer application.

2. The method of claim 1 wherein the step of capturing the stereo image further includes capturing the stereo image using a stereo camera.

3. The method of claim 1 further including recognizing a gesture associated with the object by analyzing changes in the position information of the object, and controlling the computer application based on the recognized gesture.

4. The method of claim 3 further including:

determining an application state of the computer application; and

using the application state in recognizing the gesture.

5. The method of claim 1 wherein the object is the user.

6. The method of claim 1 wherein the object is a part of the user.

7. The method of claim 1 further including providing feedback to the user relative to the computer application.

8. The method of claim 1 wherein processing the stereo image to determine position information of the object further includes mapping the position information from position coordinates associated with the object to screen coordinates associated with the computer application.

9. The method of claim 1 wherein processing the stereo image further includes processing the stereo image to identify feature information and produce a scene description from the feature information.

10. The method of claim 9 further including analyzing the scene description in a scene analysis process to determine position information of the object.

11. The method of claim 9 wherein processing the stereo image further includes:

analyzing the scene description to identify a change in position of the object; and

mapping the change in position of the object.

12. The method of claim 9 wherein processing the stereo image to produce the scene description further includes:

processing the stereo image to identify matching pairs of features in the stereo image; and

calculating a disparity and a position for each matching feature pair to create a scene description.

13. The method of claim 12 wherein:

capturing the stereo image further includes capturing a reference image from a reference camera and a comparison image from a comparison camera; and

processing the stereo image further includes processing the reference image and the comparison image to create pairs of features.

14. The method of claim 13 wherein processing the stereo image to identify matching pairs of features in the stereo image further includes:

identifying features in the reference image;

generating for each feature in the reference image a set of candidate matching features in the comparison image; and

producing a feature pair by selecting a best matching feature from the set of candidate matching features for each feature in the reference image.

15. The method of claim 13 wherein processing the stereo image further includes filtering the reference image and the comparison image.

16. The method of claim 14 wherein producing the feature pair further includes:

calculating a match score and rank for each of the candidate matching features; and

selecting the candidate matching feature with the highest match score to produce the feature pair.

17. The method of claim 14 wherein generating for each feature in the reference image, a set of candidate matching features further includes; selecting candidate matching features from a predefined range in the comparison image.

18. The method of claim 16 wherein feature pairs are eliminated based upon the match score of the candidate matching feature.

19. The method of claim 18 wherein feature pairs are eliminated if the match score of the top ranking candidate matching feature is below a predefined threshold.

20. The method of claim 18 wherein the feature pair is eliminated if the match score of the top ranking candidate matching feature is within a predefined threshold of the match score of a lower ranking candidate matching feature.

21. The method of claim 16 wherein calculating the match score further includes:

identifying those feature pairs that are neighboring;

adjusting the match score of feature pairs in proportion to the match score of neighboring candidate matching features at similar disparity; and

selecting the candidate matching feature with the highest adjusted match score to create the feature pair.

22. The method of claim 16 wherein feature pairs are eliminated by:

applying the comparison image as the reference image and the reference image as the comparison image to produce a second set of feature pairs; and

eliminating those feature pairs in the original set of feature pairs which do not have a corresponding feature pair in the second set of feature pairs.

23. The method of claim 12 further comprising:

for each feature pair in the scene description, calculating real world coordinates by transforming the disparity and position of each feature pair relative to the real world coordinates of the stereo image.

24. The method of claim 14 wherein selecting features further includes dividing the reference image and the comparison image of the stereo image into blocks.

25. The method of claim 24 wherein the feature is described by a pattern of luminance of the pixels contained with the blocks.

26. The method of claim 24 wherein dividing further includes dividing the images into pixel blocks having a fixed size.

27. The method of claim 26 wherein the pixel blocks are 8 x 8 pixel blocks.

28. The method of claim 10 wherein analyzing the scene description to determine the position information of the object further includes cropping the scene description to exclude feature information lying outside of a region of interest in a field of view.

29. The method of claim 28 wherein cropping further includes establishing a boundary of the region of interest.

30. The method of claim 10 wherein analyzing the scene description to determine the position information of the object further includes:

clustering the feature information in a region of interest into clusters having a collection of features by comparison to neighboring feature information within a predefined range; and

calculating a position for each of the clusters.

31. The method of claim 30 further including eliminating those clusters having less than a predefined threshold of features.

32. The method of claim 30 further including:

selecting the position of the clusters that match a predefined criteria;

recording the position of the clusters that match the predefined criteria as object position coordinates; and

outputting the object position coordinates.

33. The method of claim 30 further including determining the presence of a user from the clusters by checking features within a presence detection region.

34. The method of claim 32 wherein calculating the position for each of the clusters excludes those features in the clusters that are outside of an object detection region.

35. The method of claim 32 further including defining a dynamic object detection region based on the object position coordinates.

36. The method of claim 35 wherein the dynamic object detection region is defined relative to a user's body.

37. The method of claim 32 further including defining a body position detection region based on the object position coordinates.

38. The method of claim 37 wherein defining the body position detection region further includes detecting a head position of the user.

39. The method of claim 32 further including smoothing the motion of the object position coordinates to eliminate jitter between consecutive image frames.

40. The method of claim 32 further including calculating hand orientation information from the object position coordinates.

41. The method of claim 40 wherein outputting the object position coordinates further includes outputting the hand orientation information.

42. The method of claim 40 further including smoothing the changes in the hand orientation information.

43. The method of claim 36 wherein defining the dynamic object detection region includes:

identifying a position of a torso-divisioning plane from the collection of features; and

determining the position of a hand detection region relative to the torso-divisioning plane in the axis perpendicular to the torso divisioning plane.

44. The method of claim 43 further including:

identifying a body center position and a body boundary position from the collection of features;

identifying a position indicating part of an arm of the user from the collection of features using the intersection of the feature pair cluster with the torso divisioning plane; and

identifying the arm as either a left arm or a right arm using the arm position relative to the body position.

45. The method of claim 44 further including establishing a shoulder position from the body center position, the body boundary position, the torso-divisioning plane, and the left arm or the right arm identification.

46. The method of claim 45 wherein defining the dynamic object detection region includes determining position data for the hand detection region relative to the shoulder position.

47. The method of claim 46 further including smoothing the position data for the hand detection region.

48. The method of claim 45 further including:

determining the position of the dynamic object detection region relative to the torso divisioning plane in the axis perpendicular to the torso divisioning plane;

determining the position of the dynamic object detection region in the horizontal axis relative to the shoulder position; and

determining the position of the dynamic object detection region in the vertical axis relative to an overall height of the user using the body boundary position.

49. The method of claim 36 wherein defining the dynamic object detection region includes:

establishing the position of a top of the user's head using topmost feature pairs of the collection of features unless the topmost feature pairs are at the boundary; and

determining the position of a hand detection region relative to the top of the user's head.

50. A method of using stereo vision to interface with a computer, the method comprising:

capturing a stereo image using a stereo camera;

processing the stereo image to determine position information of an object in the stereo image, the object being controlled by a user;

processing the stereo image to identify feature information, to produce a scene description from the feature information, and to identify matching pairs of features in the stereo image;

calculating a disparity and a position for each matching feature pair to create the scene description;

analyzing the scene description in a scene analysis process to determine position information of the object;

clustering the feature information in a region of interest into clusters having a collection of features by comparison to neighboring feature information within a predefined range;

calculating a position for each of the clusters; and

using the position information allow the user to interact with a computer application.

51. The method of claim 50 further including:

mapping the position of the object from the feature information from camera coordinates to screen coordinates associated with the computer application; and

using the mapped position to interface with the computer application.

52. The method of claim 50 further including:

recognizing a gesture associated with the object by analyzing changes in the position information of the object in the scene description; and

combining the position information and the gesture to interface with the computer application.

53. The method of claim 50 wherein the step of capturing the stereo image further includes capturing the stereo image using a stereo camera.

54. A stereo vision system for interfacing with an application program running on a computer, the stereo vision system comprising:

first and second video cameras arranged in an adjacent configuration and operable to produce a series of stereo video images; and

a processor operable to receive the series of stereo video images and detect objects appearing in an intersecting field of view of the cameras, the processor executing a process to:

define an object detection region in three-dimensional coordinates relative to a position of the first and second video cameras;

select a control object appearing within the object detection region; and

map position coordinates of the control object to a position indicator associated with the application program as the control object moves within the object detection region.

55. The stereo vision system of claim 54 wherein the process selects as a control object a detected object appearing closest to the video cameras and within the object detection region.

56. The stereo vision system of claim 54 wherein the control object is a human hand.

57. The stereo vision system of claim 54 wherein a horizontal position of the control object relative to the video cameras is mapped to an x-axis screen coordinate of the position indicator.

58. The stereo vision system of claim 54 wherein a vertical position of the control object relative to the video cameras is mapped to a y-axis screen coordinate of the position indicator.

59. The stereo vision system of claim 54 wherein the processor is configured to:

map a horizontal position of the control object relative to the video cameras to a x-axis screen coordinate of the position indicator;

map a vertical position of the control object relative to the video cameras to a y-axis screen coordinate of the position indicator; and

emulate a mouse function using the combined x-axis and y-axis screen coordinates provided to the application program.

60. The stereo vision system of claim 59 wherein the processor is further configured to emulate buttons of a mouse using gestures derived from the motion of the object position.

61. The stereo vision system of claim 59 wherein the processor is further configured to emulate buttons of a mouse based upon a sustained position of the control object in any position within the object detection region for a predetermined time period.

62. The stereo vision system of claim 59 wherein the processor is further configured to emulate buttons of a mouse based upon a position of the position indicator being sustained within the bounds of an interactive display region for a predetermined time period.

63. The stereo vision system of claim 54 wherein the processor is further configured to map a z-axis depth position of the control object relative to the video cameras to a virtual z-axis screen coordinate of the position indicator.

64. The stereo vision system of claim 54 wherein the processor is further configured to:

map a x-axis position of the control object relative to the video cameras to an x-axis screen coordinate of the position indicator;

map a y-axis position of the control object relative to the video cameras to a y-axis screen coordinate of the position indicator; and

map a z-axis depth position of the control object relative to the video cameras to a virtual z-axis screen coordinate of the position indicator.

65 The stereo vision system of claim 64 wherein a position of the position indicator being within the bounds of an interactive display region triggers an action within the application program.

66. The stereo vision system of claim 54 wherein movement of the control object along a z-axis depth position that covers a predetermined distance within a predetermined time period triggers a selection action within the application program.

67. The stereo vision system of claim 54 wherein a position of the control object being sustained in any position within the object detection region for a predetermined time period triggers a selection action within the application program.

68. A stereo vision system for interfacing with an application program running on a computer, the stereo vision system comprising:

first and second video cameras arranged in an adjacent configuration and operable to produce a series of stereo video images; and

a processor operable to receive the series of stereo video images and detect objects appearing in the intersecting field of view of the cameras, the processor executing a process to:

define an object detection region in three-dimensional coordinates relative to a position of the first and second video cameras;

select as a control object a detected object appearing closest to the video cameras and within the object detection region;

define sub regions within the object detection region;

identify a sub region occupied by the control object;

associate with that sub region an action that is activated when the control object occupies that sub region; and

apply the action to interface with a computer application.

69. The stereo vision system of claim 68 wherein the action associated with the sub region is further defined to be an emulation of the activation of keys associated with a computer keyboard.

70. The stereo vision system of claim 68 wherein a position of the control object being sustained in any sub region for a predetermined time period triggers the action.

71. A stereo vision system for interfacing with an application program running on a computer, the stereo vision system comprising:

first and second video cameras arranged in an adjacent configuration and operable to produce a series of stereo video images; and

a processor operable to receive the series of stereo video images and detect objects appearing in an intersecting field of view of the cameras, the processor executing a process to:

identify an object perceived as the largest object appearing in the intersecting field of view of the cameras and positioned at a predetermined depth range;

select the object as an object of interest;

determine a position coordinate representing a position of the object of interest; and

use the position coordinate as an object control point to control the application program.

72. The system of claim 71 wherein the process causes the processor to:

determine and store a neutral control point position;

map a coordinate of the object control point relative to the neutral control point position; and

use the mapped object control point coordinate to control the application program.

73. The system of claim 72 wherein the process causes the processor to:

define a region having a position based upon the position of the neutral control point position;

map the object control point relative to its position within the region; and

use the mapped object control point coordinate to control the application program.

74. The system of claim 72 wherein the process causes the processor to:

transform the mapped object control point to a velocity function;

determine a viewpoint associated with a virtual environment of the application program; and

use the velocity function to move the viewpoint within the virtual environment.

75. The system of claim 71 wherein the process causes the processor to map a coordinate of the object control point to control a position of an indicator within the application program.

76. The system of claim 75 wherein the indicator is an avatar.

77. The system of claim 71 wherein the process causes the processor to map a coordinate of the object control point to control an appearance of an indicator within the application program.

78. The system of claim 77 wherein the indicator is an avatar.

79. The system of claim 71 wherein the object of interest is a human appearing within the intersecting field of view.

80. A stereo vision system for interfacing with an application program running on a computer, the stereo vision system comprising:

first and second video cameras arranged in an adjacent configuration and operable to produce a series of stereo video images; and

a processor operable to receive the series of stereo video images and detect objects appearing in an intersecting field of view of the cameras, the processor executing a process to:

identify an object perceived as the largest object appearing in the intersecting field of view of the cameras and positioned at a predetermined depth range;

select the object as an object of interest;

define a control region between the cameras and the object of interest, the control region being positioned at a predetermined location and having a predetermined size relative to a size and a location of the object of interest;

search the control region for a point associated with the object of interest that is closest to the cameras and within the control region;

select the point associated with the object of interest as a control point if the point associated with the object of interest is within the control region; and

map position coordinates of the control point, as the control point moves within the control region, to a position indicator associated with the application program.

81. The system of claim 80 wherein the processor is operable to:

map a horizontal position of the control point relative to the video cameras to an x-axis screen coordinate of the position indicator;

map a vertical position of the control point relative to the video cameras to a y-axis screen coordinate of the position indicator; and

emulate a mouse function using a combination of the x-axis and the y-axis screen coordinates.

82. The system of claim 80 wherein the processor is operable to:

map a x-axis position of the control point relative to the video cameras to an x-axis screen coordinate of the position indicator;

map a y-axis position of the control point relative to the video cameras to a y-axis screen coordinate of the position indicator; and

map a z-axis depth position of the control point relative to the video cameras to a virtual z-axis screen coordinate of the position indicator.

83. The system of claim 80 wherein the object of interest is a human appearing within the intersecting field of view.

84. The system of claim 80 wherein the control point is associated with a human hand appearing within the control region.

85. A stereo vision system for interfacing with an application program running on a computer, the stereo vision system comprising:

first and second video cameras arranged in an adjacent configuration and operable to produce a series of stereo video images; and

a processor operable to receive the series of stereo video images and detect objects appearing in an intersecting field of view of the cameras, the processor executing a process to:

define an object detection region in three-dimensional coordinates relative to a position of the first and second video cameras;

select up to two hand objects from the objects appearing in the intersecting field of view that are within the object detection region; and

map position coordinates of the hand objects, as the hand objects move within the object detection region, to positions of virtual hands associated with an avatar rendered by the application program.

86. The system of claim 85 wherein the process selects the up to two hand objects from the objects appearing in the intersecting field of view that are closest to the video cameras and within the object detection region.

87. The system of claim 85 wherein the avatar takes the form of a human-like body.

88. The system of claim 85 wherein the avatar is rendered in and interacts with a virtual environment forming part of the application program.

89. The system of claim 88 wherein the processor further executes a process to compare the positions of the virtual hands associated with the avatar to positions of virtual objects within the virtual environment to enable a user to interact with the virtual objects within the virtual environment.

90. The system of claim 85 wherein the processor further executes a process to:

detect position coordinates of a user within the intersecting field of view; and

map the position coordinates of the user to a virtual torso of the avatar rendered by the application program.

91. The system of claim 85 wherein the process moves at least one of the virtual hands associated with the avatar to a neutral position if a corresponding hand object is not selected.

92. The system of claim 85 wherein the processor further executes a process to:

detect position coordinates of a user within the intersecting field of view; and

map the position coordinates of the user to a velocity function that is applied to the avatar to enable the avatar to roam through a virtual environment rendered by the application program.

93. The system of claim 92 wherein the velocity function includes a neutral position denoting zero velocity of the avatar.

94. The system of claim 93 wherein the processor further executes a process to map the position coordinates of the user relative to the neutral position into torso coordinates associated with the avatar so that the avatar appears to lean.

95. The system of claim 92 wherein the processor further executed a process to compare the position of the virtual hands associated with the avatar to positions of virtual objects within the virtual environment to enable the user to interact with the virtual objects while roaming through the virtual environment.

96. The system of claim 85 wherein a virtual knee position associated with the avatar is derived by the application program and used to refine an appearance of the avatar.

97. The system of claim 85 wherein a virtual elbow position associated with the avatar is derived by the application program and used to refine an appearance of the avatar.

98. The system of claim 85 further comprising a third video camera arranged in an adjacent configuration with the first and second video cameras and operable to produce the series of stereo video images.